

Theory and Practice of Data Assimilation in Ocean Modeling

Robert N. Miller

College of Oceanic and Atmospheric Sciences

Oregon State University

Oceanography Admin. Bldg. 104

Corvallis, OR 97331-5503

Phone: (541) 737-4555 Fax: (541) 737-2064 Email: miller@coas.oregonstate.edu

Award Number: N00014-04-1-0262

LONG-TERM GOALS

The long-range goal of this project is to form the best picture of the ocean as an evolving system based on data assimilation, i.e, the construction of a composite estimate of the state of the ocean based on a combination of observed data with computational model output, and to use that picture to understand the physical processes that govern the ocean's behavior. Oceanic observations are sparse and models are limited in accuracy, but taken together, one can form a quantitative description of the state of the ocean that is superior to any based on either models or data alone. Along with the goals of analysis and prediction, we seek reliable estimates of the errors in our results. We expect our results to have implications beyond the technical challenges of data assimilation. In particular, we believe this research will lead to enhanced understanding of the implications of nonlinearity and randomness for predictability of the ocean and atmosphere.

In keeping with our goal of providing reliable error estimates for our data assimilation products, we seek to develop efficient methods for estimating useful statistical measures of errors in stochastic forecast models, and information about stochastic systems is contained in the associated probability density function (PDF). The PDFs of nonlinear stochastic models are not, in general, Gaussian, so we must find methods for forecast evaluation based on information about the particular PDF generated by the model.

Since our goal is the development of practical analysis and forecast systems for the ocean, we want to solve remaining scientific problems involved in transition from data assimilation experiments tuned to specific models and data sets to operational analysis and prediction on a research basis. This will involve rigorous quantification of the information content of each data set, as well as quality control, a problem with which the ocean modeling community has limited experience.

OBJECTIVES

The principal objective of this project is the development, implementation and evaluation of practical data assimilation methods for regional to basin scale ocean models. Since data assimilation methods that give the most and best information are highly resource intensive, and often not practical for use with detailed models, we are particularly interested in the price paid in terms of accuracy and confidence for using economical but suboptimal data assimilation methods.

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 30 SEP 2008		2. REPORT TYPE Annual		3. DATES COVERED 00-00-2008 to 00-00-2008	
4. TITLE AND SUBTITLE Theory And Practice Of Data Assimilation In Ocean Modeling				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Oregon State University, College of Oceanic and Atmospheric Sciences, Oceanography Admin. Bldg. 104, Corvallis, OR, 97331-5503				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES code 1 only					
14. ABSTRACT The long-range goal of this project is to form the best picture of the ocean as an evolving system based on data assimilation, i.e, the construction of a composite estimate of the state of the ocean based on a combination of observed data with computational model output, and to use that picture to understand the physical processes that govern the ocean's behavior. Oceanic observations are sparse and models are limited in accuracy, but taken together, one can form a quantitative description of the state of the ocean that is superior to any based on either models or data alone. Along with the goals of analysis and prediction, we seek reliable estimates of the errors in our results. We expect our results to have implications beyond the technical challenges of data assimilation. In particular, we believe this research will lead to enhanced understanding of the implications of nonlinearity and randomness for predictability of the ocean and atmosphere.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 6	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Direct calculation of full PDFs is not feasible for practical models of the ocean or atmosphere, but useful approximations to the PDF can be calculated from Monte-Carlo experiments, by virtue of the fact that the number of truly independent degrees of freedom in practical models is very much smaller than the dimension of the state vector. This intuition is the motivation for the ensemble methods that have become popular in recent years.

Our experience with Monte-Carlo methods in simplified systems has led us to investigate the details of methods for ensemble generation that have been presented in the community. The motivation for these specialized methods for generating ensembles is precisely the specification of the PDF of a complex model whose behavior is believed to be captured by a relatively small number of independent degrees of freedom. By detailed study of the behavior of ensembles in increasingly complex models, we hope to gain the insights necessary to generate the most efficient ensembles, which should, in turn, lead to the error estimates necessary for data assimilation systems and prior estimates of forecast accuracy.

Optimized methods require accurate knowledge of the statistics of the errors in the model and the data. It is therefore an objective to understand in detail the sensitivity of the data assimilation scheme to the details of the defining error estimates.

APPROACH

The basic assumptions underlying data assimilation methods in use or proposed are known to be false to some degree. We plan to study the consequences of these assumptions by constructing a hierarchy of schemes with decreasing reliance on ad hoc assumptions. It is our guiding philosophy that the best way to learn how to design and implement the most economical methods that meet our needs is to begin by implementing methods which are as close to optimal as possible. From that point, we can quantify the loss of accuracy and the saving of resources associated with each simplification of the model or the data assimilation scheme.

Work is proceeding toward a theoretical basis for the next generation of data assimilation methods in which randomness and nonlinearity must be taken into account. To this end, we are applying tools from stochastic differential equations and from dynamical systems theory. Since our model systems are characterized by high dimensional state spaces, Monte Carlo methods must be used to study the behavior of the stochastic systems.

The theory of nonlinear filtering provides a framework in which problems of data assimilation with nonlinear models and non-Gaussian noise sources can be treated (see, e.g., Miller et al., 1999). In the case of linear models and Gaussian noise sources, this theory reduces to the familiar Kalman filter. In the formal theory of nonlinear filtering, the final result is not a single model state vector or trajectory in state space, but a PDF defined as a scalar function of the state variables and time. From this PDF, the mean, median, mode, or other statistic can be computed for use as the working estimate of the state of the system, along with the desired confidence intervals. The assignment of confidence limits corresponds in the case of a group of particles in physical space to drawing contours in the spatial domain which can be expected to define a region which contains, say, 90% of the particles.

The problem is that for even schematic models of the ocean or atmosphere, an unrealistically large number of particle trajectories in phase space must be calculated in order to represent the PDF faithfully. Useful ensemble analysis therefore requires judicious choice of ensemble members. We have concentrated our recent efforts on evaluation of ensemble methods, which we see as facilitating

the generation of the forecast error estimates necessary for data assimilation. These forecast error estimates are of interest in and of themselves, since they have the potential of providing a priori estimates of the reliability of a given forecast.

Results from the theory of dynamical systems lead to methods for explicit construction of the low dimensional spaces in which meaningful probabilistic calculations can be performed on complex systems. We are now finishing our work on a local model of the Kuroshio, and have begun to extend it to a model of the Pacific basin. The simplest of our models is a regional two-layer quasigeostrophic model that reproduces the observed bimodality. It operates on a state space with several thousand dimensions. This is two orders of magnitude greater than that of earlier schematic models, and, for this reason alone, presents significant technical challenges.

We now have a basis of comparison with more complex models, up to and including eddy resolving primitive equation models of the north Pacific. We are now in the process of applying our methods from dynamical systems and stochastic calculus to a suite of models, in order to understand propagation of errors and the evolution of the PDF arising from random initial and boundary conditions in a state space of workable dimension. This should allow us to construct reliable data assimilation systems for use with simulated and real data from the Kuroshio.

Many different models, based on fundamentally different physical assumptions, exhibit the observed bimodality of the Kuroshio in some form. We are now in the process of comparing our model to different models and to observed data in order to determine a basis for distinction between the physical mechanisms in the different models.

Technical support for this project is provided by Ms. Laura Ehret.

WORK COMPLETED

We have completed our suite of models with the addition of a new model of the Pacific basin with eddy-resolving horizontal resolution and coarse vertical resolution. This model is suitable for efficient implementation of advanced data assimilation methods, and produces results that are directly comparable to our regional model and to existing detailed general circulation models.

We have produced a state transition in the model and established a correspondence with steady states on our bifurcation diagram. See figures 2 and 3.

RESULTS

The divergence field from a $1/10^\circ$ model of the north Pacific (Model output kindly provided by Drs. J. McClean and M. Maltrud) is of order 0.1, appropriately scaled, the same order of magnitude as the Rossby number. This is true for the region off the coast of Japan where the Kuroshio is strong; see figure 1.

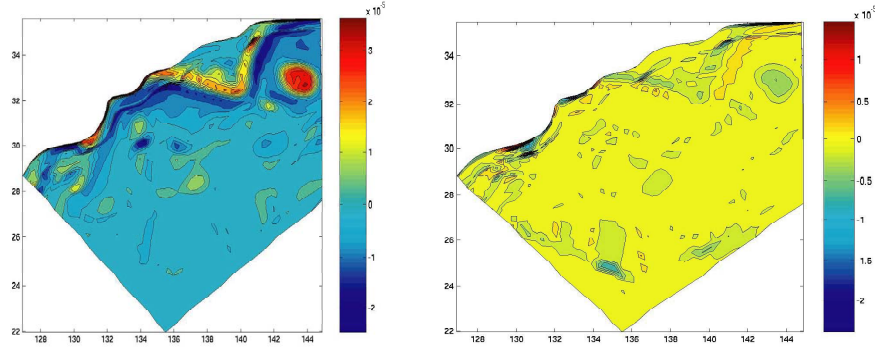


Figure 1. Relative vorticity (left) and divergence (right) fields derived from a $1/10^\circ$ simulation of the north Pacific off the coast of Japan (see Smith et al., 2000). Note the signature of the Kuroshio in the vorticity field, and the expected small scale features in the divergence.

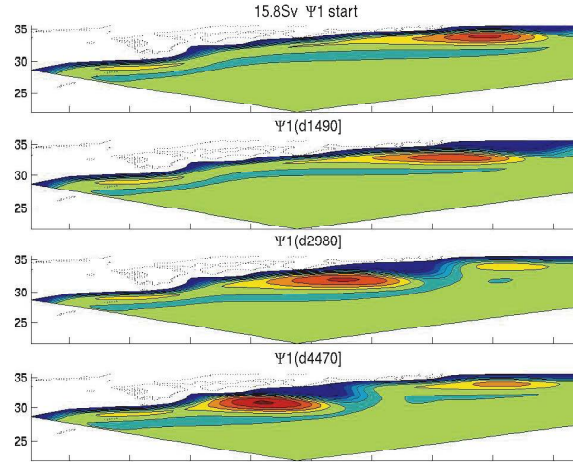


Figure 2. Transition from small to large meander steady states in the quasigeostrophic model

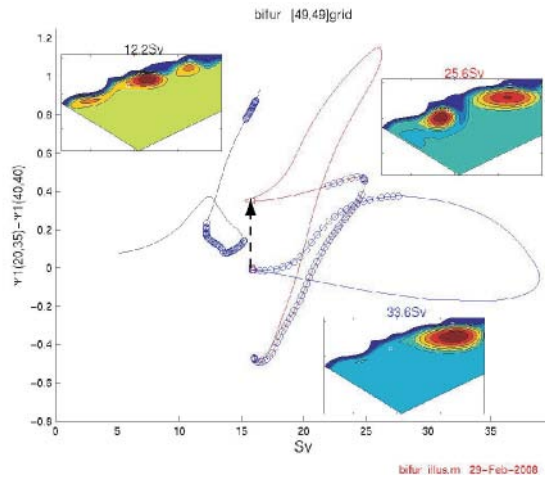


Figure 3. *Bifurcation diagram of the quasigeostrophic model. Dashed line with arrow indicating direction depicts the path of the transition depicted in figure 1.*

IMPACT/APPLICATIONS

Major weather centers, including the US National Center for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasting (ECMWF) now use ensemble methods for operational forecast validation; see Molteni et al. (1996), Toth and Kalnay (1993). Our work on Monte-Carlo methods should provide enhanced capability for validation of forecasts of the ocean and atmosphere, in addition to application to data assimilation. Our work on breeding modes and planned work on other schemes for ensemble generation should provide significant guidance in optimizing methods for generation of ensembles. Our work on dynamical analysis of models of the Kuroshio should lead to practical methods for identification of low-dimensional spaces in which efficient ensemble methods could be implemented.

We expect our the results of our inverse model of the Kuroshio to shed light on the importance of nonlinearity in ocean models; further, we expect that our work with comparisons among models and data for the Kuroshio will lead to greater insight into the intrinsic variability of basin-scale ocean circulation.

RELATED PROJECTS

Estimating the representation error of satellite and in-situ data for data assimilation into ocean models. Work is in progress on estimation of representation error in the ocean component of the NCEP climate model.

REFERENCES

- Miller, R. N., E. F. Carter and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167-194.
- Molteni, F., R. Buizza, T. N. Palmer and T. Petroligis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73-119.

Smith, R. D., Maltrud, M. E., Bryan, F. O. and Hecht, M. W., 2000: Numerical simulation of the north Atlantic at $1/10^\circ$. *J. Phys. Oceanogr.* 30, 1532-1560.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317-2330.

PUBLICATIONS

Miller, R. N. (2007) Topics in Data Assimilation : Stochastic Processes. *Physica D*, **230**, 17-26 doi:10.1016/j.physd.2006.07.015 [published, refereed]

Miller, R. N. (2007) Discussion: Subjective likelihood for the assessment of trends in the ocean's mixed layer depth. *Journal of the American Statistical Association.*, 102, 781-782. [published, refereed]